様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ５ 月 ３０ 日現在

機関番号：６２６１５
研究種目：基盤研究(B)（一般）
研究期間：2021～2023
課題番号：２１Ｈ０３４１９

研究課題名（和文）更新を対象としたデータ相互運用問題のソフトウェア基盤技術

研究課題名（英文）The software infrastructure technology for data interoperability focusing on updates

研究代表者
加藤　弘之（Kato, Hiroyuki）

国立情報学研究所・アーキテクチャ科学研究系・助教

研究者番号：１０３２１５８０

交付決定額（研究期間全体）：（直接経費）　　10,550,000円

研究成果の概要（和文）：更新を取り扱うことができるようなデータ相互運用問題のソフトウェア基盤技術として、双方向変換を用いた。P2Pアプローチにおけるスキーマの変更やデータの変更は、双方向変換の２部グラフ構成による同期網における節・枝の進化・縮退による相互運用条件の適応に対応づけた。
また、双方向変換を構成する順方向変換が関係代数演算の選択または射影に対応し、逆方向変換が単調性と最小性を満たすような双方向変換を対象として、それが与えられた関数従属性に対して整合性をもつための判定可能な必要十分条件を与えた。

研究成果の学術的意義や社会的意義
データ相互運用問題で更新が扱えるようになるので、同じ意味を持つ異なるフォーマットで存在しているデータ全体を扱った更新が可能となる。特に、既に稼働しているアプリケーションが使っているデータベースを維持したまま、新しい要求に沿ったアプリケーションを開発可能となるなど、これまでは実現できなかった開発コストの削減に繋がる基盤技術に役に立つと思われる。

研究成果の概要（英文）：We utilized bidirectional transformations as the foundational technology for software capable of handling updates in data interoperability problems. In a P2P approach, schema changes and data modifications are mapped to the adaptation of interoperability conditions through the evolution and contraction of nodes and edges in a synchronization network composed of a bipartite graph structure of bidirectional transformations.
Additionally, focusing on bidirectional transformations where the forward transformation corresponds to selection or projection in relational algebra operations and the backward transformation satisfies monotonicity and minimality, we provided necessary and sufficient conditions that are decidable for ensuring consistency with given functional dependencies.

研究分野：データベースプログラミング言語

キーワード：データ相互運用問題　双方向変換　ビュー更新問題　Datalog

様　式　Ｃ－１９、Ｆ－１９－１（共通）

１．研究開始当初の背景

「データ相互運用問題」とは、独立に構築された様々なデータベース上において、異なるスキーマのもとに存在している同じ意味を表すデータを統一的に扱う問題である。データ相互運用問題は、データベースの研究分野で古くから継続して取り組まれてきた課題であり、新しい技術や応用とともに発展し続けている。実際、SIGMOD/PODS、VLDB、CIDR などのデータベース研究分野のトップの国際会議のセッションや併設ワークショップで毎年この問題が取り上げられている。例えば、VLDB2020 の併設ワークショップの一つはデータ相互運用問題に関するワークショップであり、研究代表者はこのワークショップの Co-Chairs の一人である。また、近年その重要性が指摘されているデータ科学においてもデータ相互運用問題は重要な役割を果たしている。2016 年のフォーブス誌によると、データサイエンティストの仕事のおよそ 80%は「データを用意すること」に宛てられている。ここでいう「データの用意」とは、データ分析のために、データを探し、集め、整理、統合し、管理することである。また、ACM の学会誌である CACM の 2022 年 8 月号によると、データサイエンティストは、彼らの時間の 80-90%をデータクリーニング、データ統合、データ変換に費やしていると繰り返し述べている。このような問題は、データ相互運用問題で使われている技術を使うことで軽減できる。

データ相互運用問題は、近年次の二つの点で新たな方向に拡張されている。１）統合されたデータへのアクセスについて従来は参照だけであったが、更新を対象とする試みがある。更新を対象とすることで、複数のピアが協調してある応用開発を推進することが可能となる。２）異なるスキーマの統合手法は、データ全体をカバーするスキーマによる統合から、各ピア間のスキーマの違いを変換で記述するスキーママッピングを用いた P2P 手法へと移行している。尚、データ相互運用問題では閉包性を達成するために、使われているデータモデルの問合せ言語を用いてスキーママッピングを記述する。本研究では、関係モデルを対象とするので、スキーママッピングは SQL と等価な Datalog を対象としている。P2P アプローチは、全体として統合スキーマを設計する必要が無いため、現実の問題に適している。問合せは自分がよく知っているピアに対してなされ、そのピアにスキーママッピングで繋がっている全てのピアからデータを検索することが可能となる。もう一つの利点は、ピアの参加離脱に関するモジュール性にある。新たに参加するピアは、最も似ているスキーマを持ち熟知しているピアとの間にスキーママッピングを記述すれば良いし、ピアが離脱する場合、自分と繋がっているスキーママッピングを削除するだけで良い。このように P2P に基づくアプローチは、各ピアの自律性を維持したままデータ統合が実現される利点がある。

データ相互運用問題における上記の新たな二つの拡張に対してはそれぞれ課題があることが知られている。まず、更新を考慮に入れた P2P に基づくデータ相互運用問題へのアプローチにおいて、自律した各ピアにおけるデータやスキーマの変更が与える影響はこれまであまり研究されてきていない。その本質的な理由は、（課題 A：）あるピアでの変更がスキーママッピングで繋がっている他のピアを経由して自分自身へ影響を及ぼす副作用の解析が困難であるからである。次に、各ピアの自律性を維持しながらデータ統合を行う場合の課題として、（課題 B：）互いに矛盾するデータの管理手法の確立が挙げられる。（課題 A）の副作用解析は、データベースビュー更新問題に帰着できる。ビュー更新問題とは、データベースから問い合わせによって抽出されたビューに対する更新をデータベースに反映する問題で、データベースの分野で古くから長年取り組まれていながら、解決できていない問題である。この問題の本質は、ビュー定義は単射ではないためにビューに対する更新をデータベースに反映する手法が複数存在する点にある。データ統合におけるスキーママッピングは問合せ言語で記述されているため、ピア１からピア２へのスキーママッピングが存在する場合、ピア２をピア１のビューとみなすことができる。このとき、ピア２の更新をピア１に伝播させるのは、まさにビュー更新問題となる。ビュー更新問題では、ビューの更新をデータベースに反映させた結果を用いて、もう一度ビューを計算した場合に副作用がないことがこれまで研究されている。研究代表者は最近このビュー更新問題に対して、これまでにない言語的なアプローチで取り組み、ビュー更新戦略を記述する言語を提案し、ビュー更新を自動的に検証することにより解決策を与えた。研究代表者によるこの成果は、データベース研究分野のトップの国際会議の一つである VLDB2020 で発表し、高い評価を得ている。　次に，(課題 B)の矛盾するデータ管理については、各ピアは自律しているため、他のピアに存在している同じ意味を表すデータの受理・拒否の管理が必要となる。特に理論や実験手法などが未だに確立されていないようなデータ科学の最先端においては、あるピアは他のピアと協調しながらも自分固有の理論や実験手法を確立していくことで、データ科学が発展している。このような、協同しながらデータの個人化(Collaborative Data Personalization) をするための基盤技術は確立されていない。

２．研究の目的

本研究の目的は、各ピアの自律性を維持しながら、更新を考慮したデータ相互運用のソフトウェア基盤技術を開発することである。

３．研究の方法
上記の課題に対して次の目標を設定して研究を進めた。
目標 A：副作用の無いスキーママッピングの定義静的解析に基づき，更新による副作用の無いスキーママッピングを定義する。この際、ビュー更新問題に対する申請者の最近の成果を要素技術として適用することで、この目標を達成する。
目標 B：互いに矛盾するデータの管理手法の確立ピア同士が互いに矛盾するデータを有している場合、各ピアは自分の判断で他のピアのデータの受理・拒否を制御することを目標とする。この時、実用的観点から、そのデータの来歴情報に基づくデータの制御手法の確立を目指す。データの来歴情報は、これまではどのデータを使ったかに主眼が置かれていたが、本研究ではどのピアが更新したかの更新情報も含めて拡張した来歴情報を定義する。
目標 C： 更新伝播の最適化実用的な観点から、データの伝播は高速に行われるようにする。そのための最適化機構を確立する．特に、スキーママッピングによって繋がっている複数の独立したデータベースに対する更新伝播において、更新言語の静的解析に基づく最適化を開発する。これまで、一つのデータベースに対する問合せ最適化や更新最適化は長年にわたり研究されているが、本研究の対象のように複数の独立したデータベースを対象とした更新伝播の最適化は、あまり研究されていない。

４．研究成果

（１）更新伝播による副作用の問題について以下の取り組みを行なった。まず、これまでの成果である「ビュー更新問題の解決法」は、更新された結果の状態に基づいたものであったが、これをデータに対する更新操作に基づくものへと拡張を行なった。具体的には、これまでの成果である、有効性が確認された「状態に基づく解決手法」は論理型言語 Datalog で記述されているが、これを仕様としてそこから「更新操作に基づく」Datalog 式への正しい導出手法を開発した。更に、データベースに対する更新操作は SQL を用いてなされるので、SQL の INSERT 文，DELETE 文，UPDATE 文を delta-Datalog 式に正しく変換する手法を開発した。
（２）入力された SQL の更新文を Delta-Datalog に書き換えた際の最適化技術を開発した。
また、基底表と導出表を節とし、基底表から導出表への非対称型双方向変換を枝とする 2 部グラフ構造による同期網における節･枝の進化･縮退による相互運用条件変更への適応方式において、節点における値域の相違の、スパン・余スパンによる調停方法と、変換の両端の制約条件の双方向伝播法について Relational Lens と least change に基づく型レベルのラウンドトリップ性を満たす双方向変換として予備的検討を行った。接点における更新競合については、木モデルを仮定した 3 者の競合の操作変換による解決を検討した。
（３）仕様として与えられる「状態変化」中に定義されている制約部分について、Datalog 式で記述することで、入力 SQL 更新文が変換された Delta-Datalog 式と統一的に扱う枠組みを開発した。具体的には、従来の制約記述手法として知られている negative constraints は Datalog 式で直接扱うことは出来ないの、positive constraints に変換し、引数を持たない Datalog 式に変換することで対応した。この変換によって、「更新するデータ集合の中に制約を満たさないデータが存在する場合、何も更新しない」という、標準的な更新操作に対応することも可能となった。
（４）双方向変換の 2 部グラフ構成による同期網における節・枝の進化・縮退による相互運用条件変更への適応方式について、(余)スパンを進化・縮退単位とし Bohannon 等の Relational Lenses の型システムを用いて精緻化し、構成を保存したまま制約条件の変化を型レベルで伝播させる方法を、静的双方向変換として選択、結合、射影、およびその合成の型推論に基づき提案した。
（５）順方向変換が関係代数演算の選択または射影に対応し、逆方向変換が単調性と最小性を満たすような双方向変換を対象として、それが与えられた関数従属性に対して整合性をもつための判定可能な必要十分条件を与えた。さらに、逆方向変換が関数従属性を最も満たしやすいように振る舞う双方向変換に対しても、整合性をもつための判定可能な必要十分条件を与えた。

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　0件）

１．著者名 Asano Yasuhito、Cao Yang、Hidaka Soichiro、Hu Zhenjiang、Ishihara Yasunori、Kato Hiroyuki、Nakano Keisuke、Onizuka Makoto、Sasaki Yuya、Shimizu Toshiyuki、Takeichi Masato、Xiao Chuan、Yoshikawa Masatoshi	４．巻 1457 CCIS
２．論文標題 Bidirectional Collaborative Frameworks for Decentralized Data Management	５．発行年 2022年
３．雑誌名 Communications in Computer and Information Science	６．最初と最後の頁 13～51
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-93849-9_2	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	日高　宗一郎 (Hidaka Soichiro) (70321578)	法政大学・情報科学部・教授 (32675)	
研究分担者	石原　靖哲 (Ishihara Yasunori) (00263434)	南山大学・理工学部・教授 (33917)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関